



APPLICATIONS FOR A
SOFTWARE-DEFINED ENTERPRISE

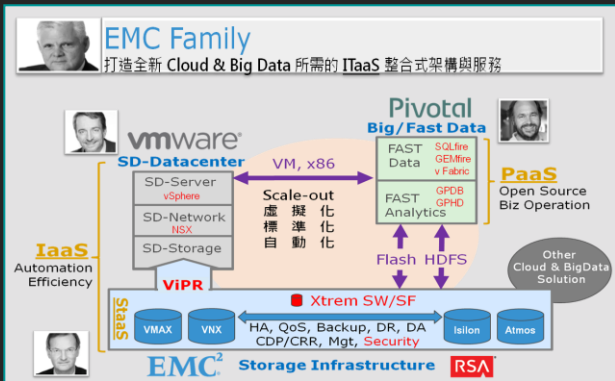
ANALYTICS FOR SPOTTING
OPPORTUNITIES

SCALE-OUT STORAGE FOR
EXPLODING DATA VOLUMES

一次搞懂 Data Technology

台灣EMC 業務拓展總監 李百飛

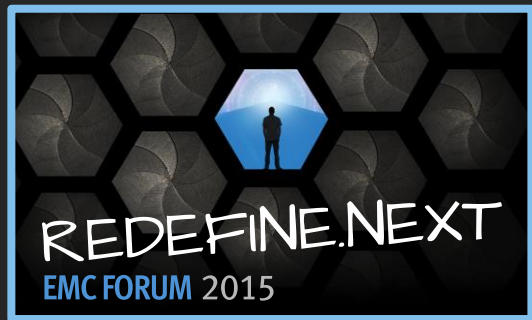
EMC引領雲端與大數據技術持續領先全球



12年中投入約**420億美元**於研發及收購
以技術領先市場

2003-2014
併購 213億美元
研發 206億美元

2015



數位轉型.未來重新定義+
(Information Generation)

2014



Mega Trend.重新定義
(Mobile . Cloud . Social . Big Data)

2013



EMC協助並引領您轉型

2012



企業和個人利用雲與大數據轉型

2011



雲與大數據交匯產生新契機

2010

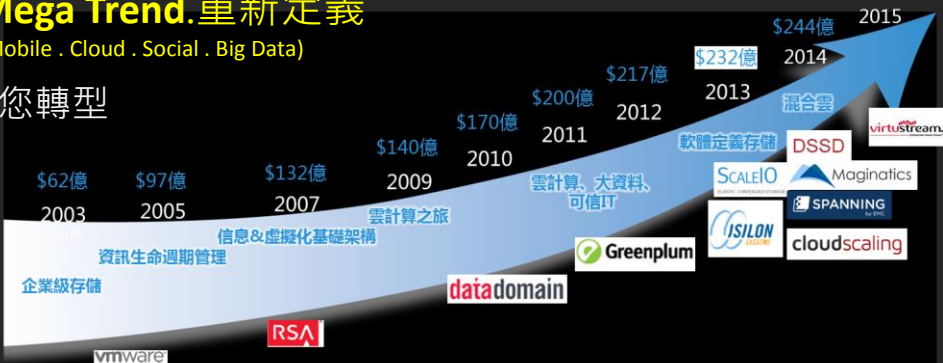


進入私有雲的旅程

2009



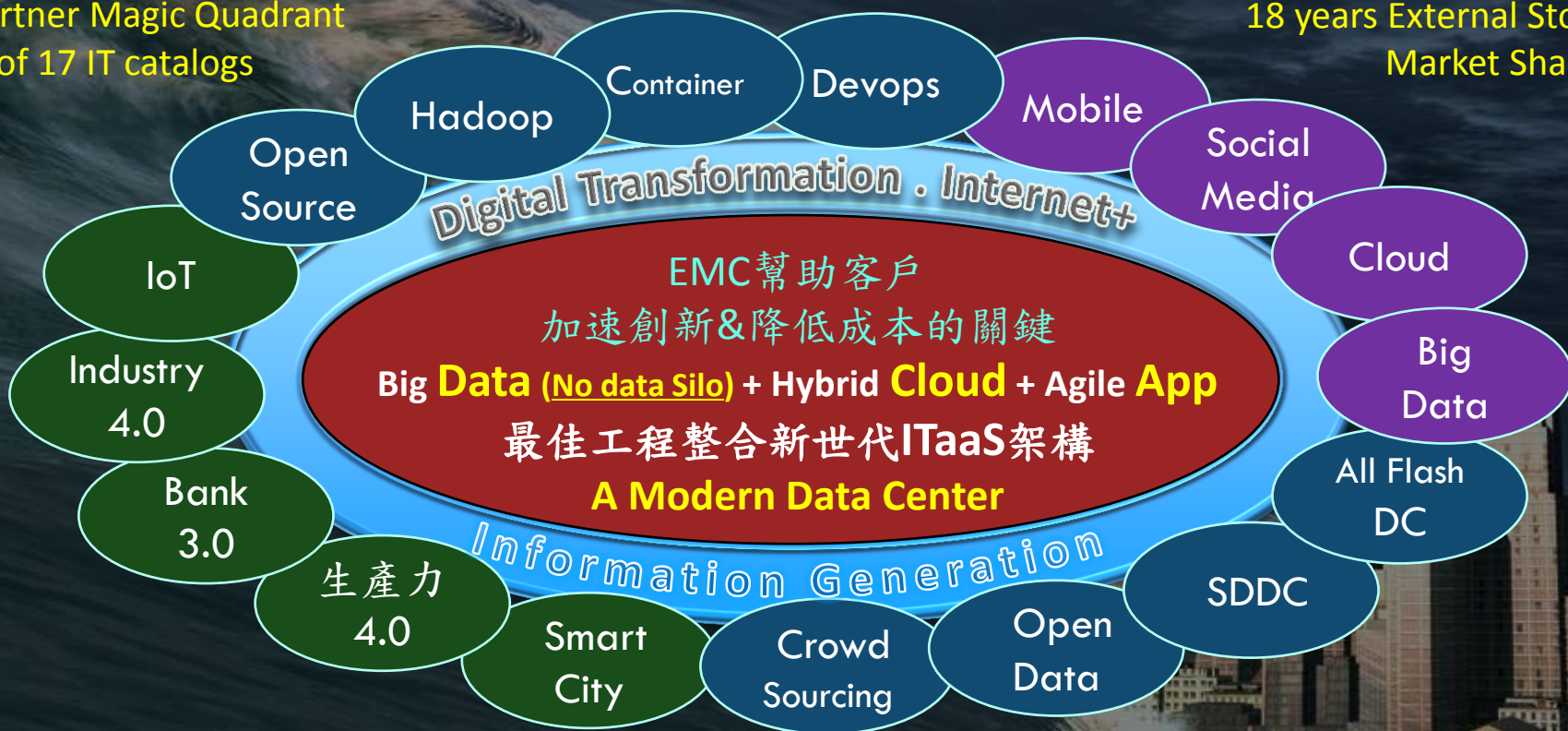
私有雲的願景



嶄新且變化快速的資訊世代已經來臨

☑ EMC is ranked "Leader"
by Gartner Magic Quadrant
in 13 of 17 IT catalogs

☑ Worldwide Continuously
18 years External Storage
Market Share #1



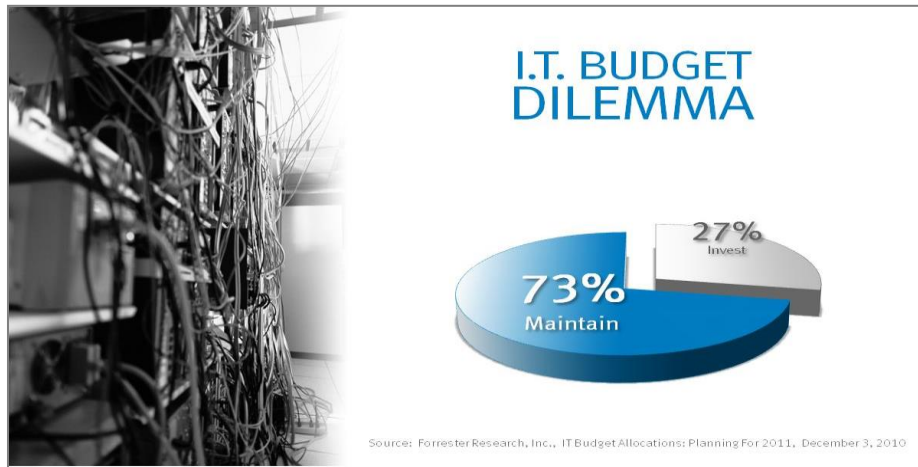
今日IT的挑戰



+

SLA UP

Business Agility



Big Data應用一個棘手的問題

資料散落在各處 (Data Silo)

- 各資料系統獨立運作
- 現有 IT 架構很難或不易整合
- 共享效率奇差無比
- 沒有全貌性的資料視野
- 高昂的系統升級成本



©Ron Leishman * illustrationsOf.com/440280

Big Data時代需要建構以資料為中心的運算架構

DATA
IS THE NEW
CENTER OF GRAVITY



Data Type: Structured, Semi-Structured, Unstructured

傳統應用環境



File Shares



HPC



Backup/Archive



1.25x



2x



2x

新型態應用環境



Analytics



Mobile



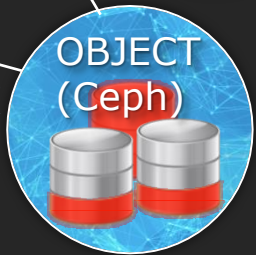
Cloud Apps



3x



3x



3x

×企業內到處
都是資料孤島
×複雜的資料流
×資料重複儲存

TRADITIONAL WORKLOADS



File Shares



HPC



Backup/Archive



NAS



SAN



TAI

NEXT-GEN WORKLOADS



DAS



CLOUD
(rest API)



OBJECT
(Ceph)



Analytics



Mobile



Cloud Apps

EMC Isilon
Data Lake

加速創新關鍵: Isilon 整合資料各種存取方法

完全解決資料孤島、複雜資料流、資料重複儲存問題



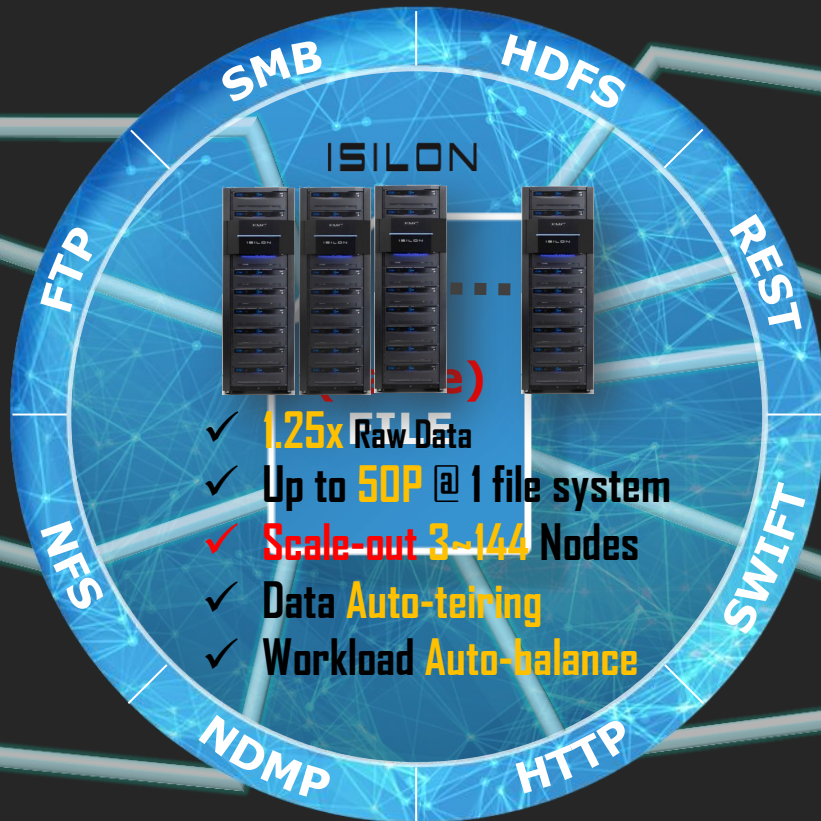
File Shares



HPC



Backup/Archive



Analytics

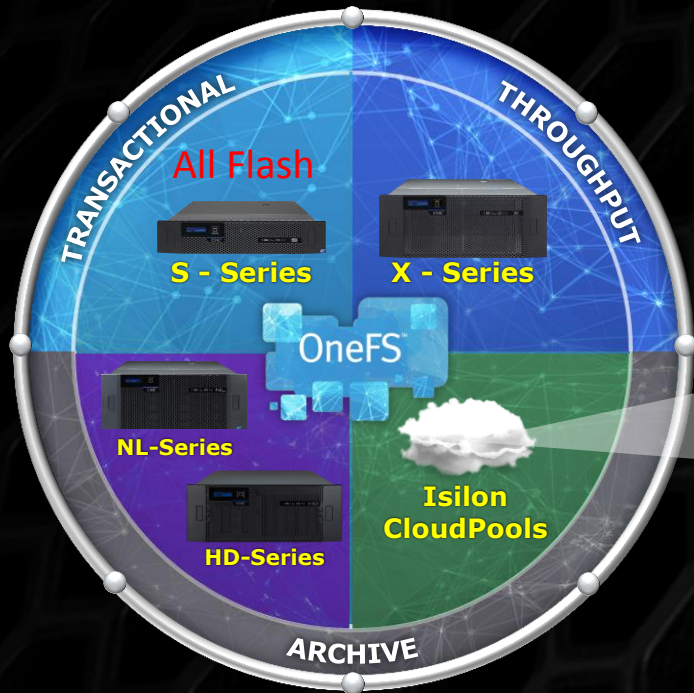


Mobile



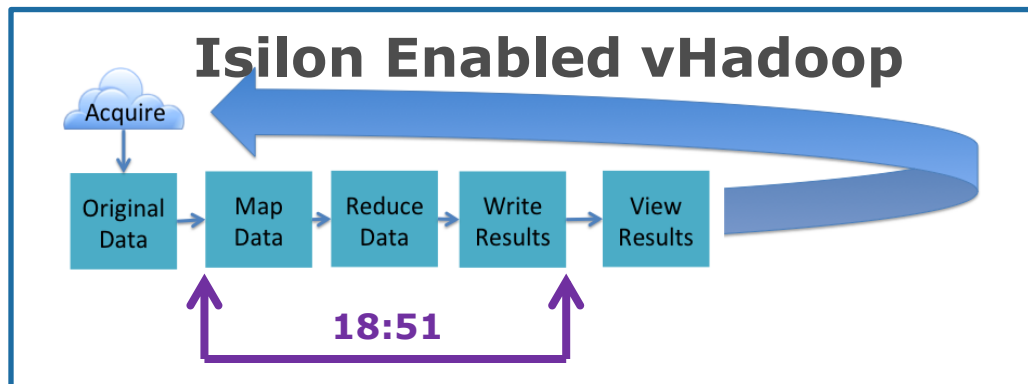
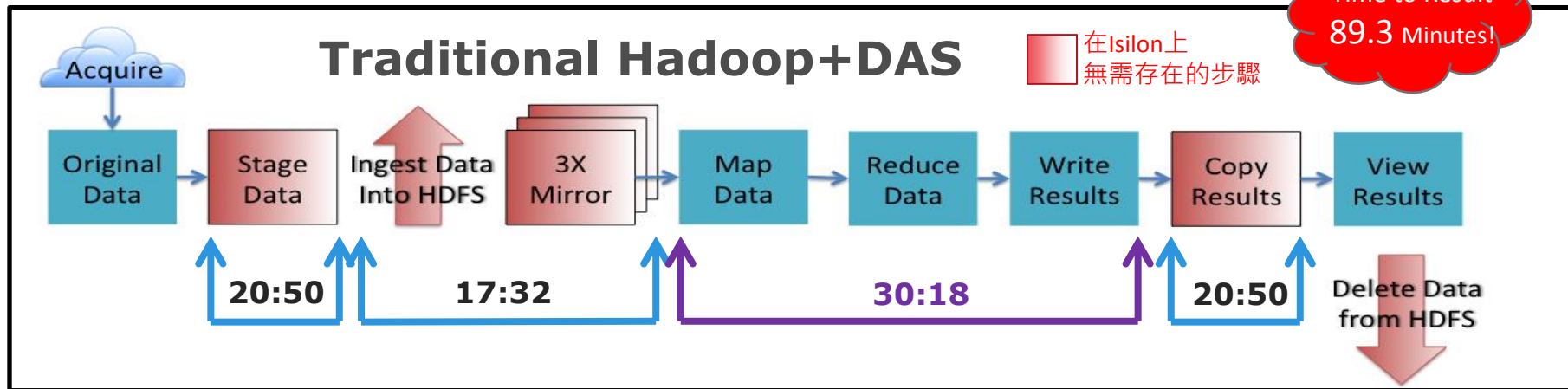
Cloud Apps

Isilon創新: All Flash + Auto tiering + Cloud-Enabled



案例: A Global Telecom 1TB Hadoop Job Cycle比較

Isilon Significantly Reduces Time To Results



Isilon Advantages

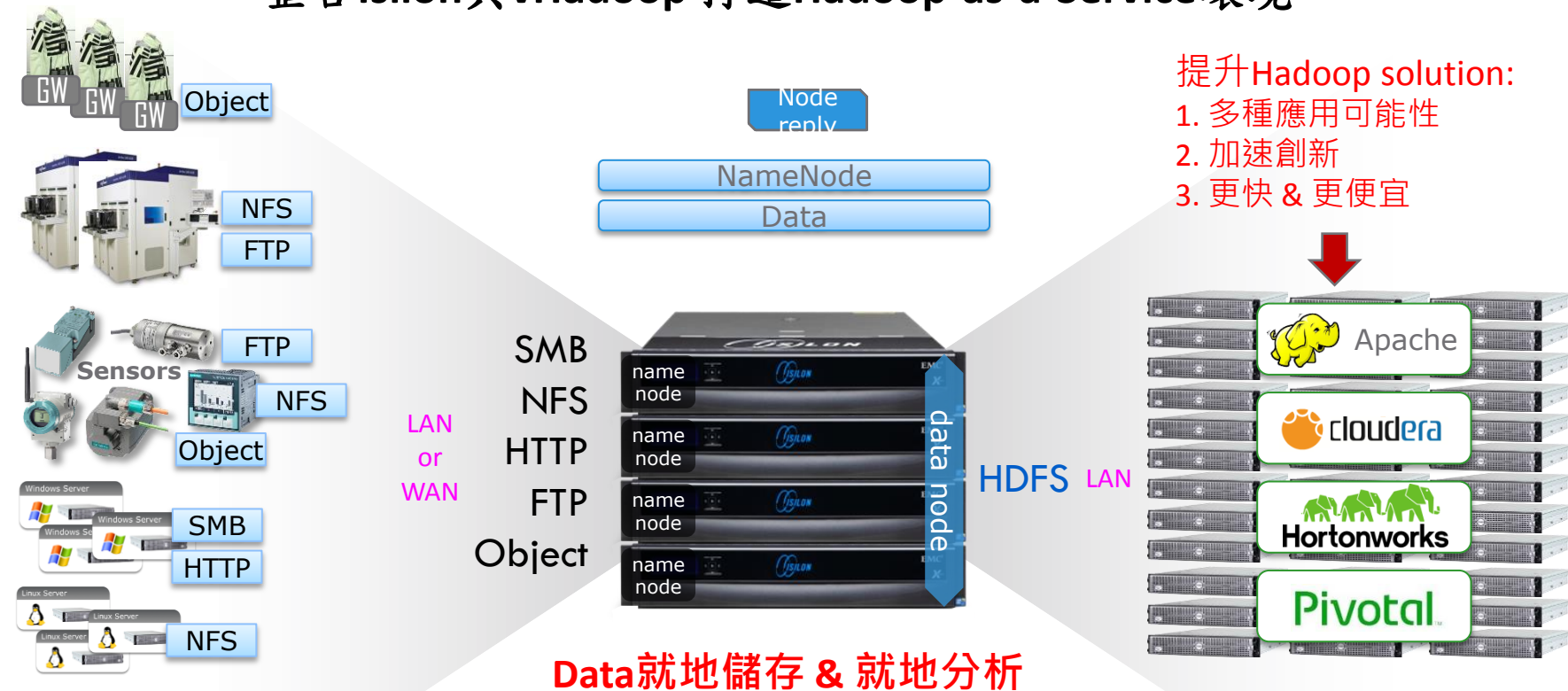
- Eliminates *All* Data Movement
- Allows for *Virtualized* Compute
- Significantly *Less* Cost
- 79% *Faster* TTR!

Terasort Test on 1TB

	DAS	Isilon	Benefit
MB/s Per Node	55.00	85.00	快55%
運算時間 (Min)	30.18	18.51	快39%
Time to Result (Min)	89.30	18.51	快79%

加速創新關鍵: 具備多種研發應用的可能性與彈性

整合Isilon與vHadoop 打造Hadoop-as-a-Service環境



提升Hadoop solution:

1. 多種應用可能性
2. 加速創新
3. 更快 & 更便宜

Data就地儲存 & 就地分析

Hadoop採先進Isilon配置與傳統DAS配置的比較分析

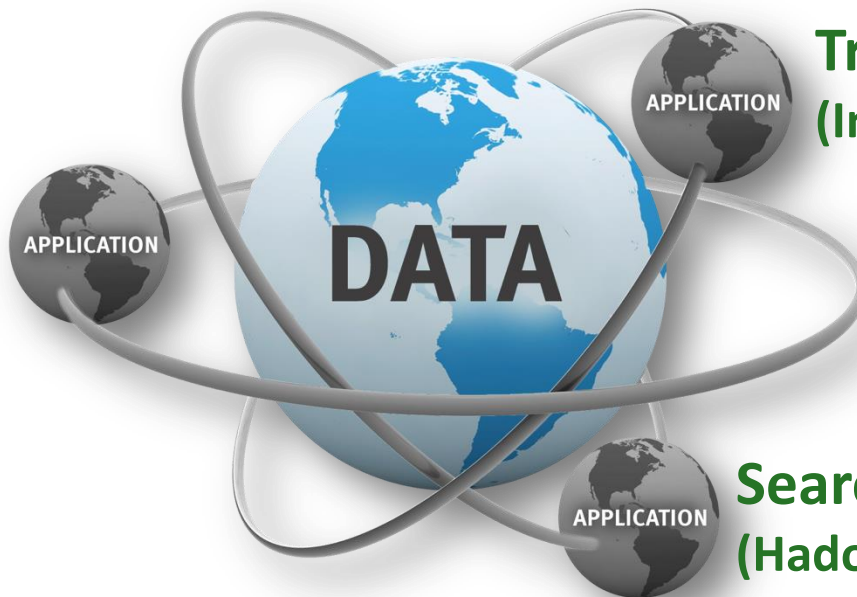
	DAS	Isilon
Data-Set Management	Ingest	In-place
Data Type	Files	Files
Protection overhead	200%	20%
NameNode Redundancy	Active/Passive	N-to-N Active
De-duplication	✗	✓
Ability to edit files/objects	✗	✓
NFS v3	✗	✓
NFS v4	✗	✓
SMB 1	✗	✓
SMB 2x	✗	✓
HTTP	✗	✓
FTP	✗	✓
Object (Proprietary)	✗	✓
HDFS v1	✗	✓
HDFS v2	✓	✓

	DAS	Isilon
Simultaneous Multi-Protocol	✗	✓
Simultaneous Hadoop Distributions	✗	✓
File/Object Level Access Control Lists	✗	✓
Snapshots	✓	✓
WORM (SEC 17a-4)	✗	✓
POSIX Compliance	✗	✓
Independent Scaling	✗	✓
Hadoop Distribution Portability	✗	✓
HAWQ Support	✓	✓
Encryption	✗	SED
Data Tiering	✗	✓
Hadoop Distributions	1	All
Consolidated Hadoop Management	✓	✗
Disaster Recovery	Full File Copy	Snap

Big Data有那些應用類型?



**Analytics-based
(EDW)**



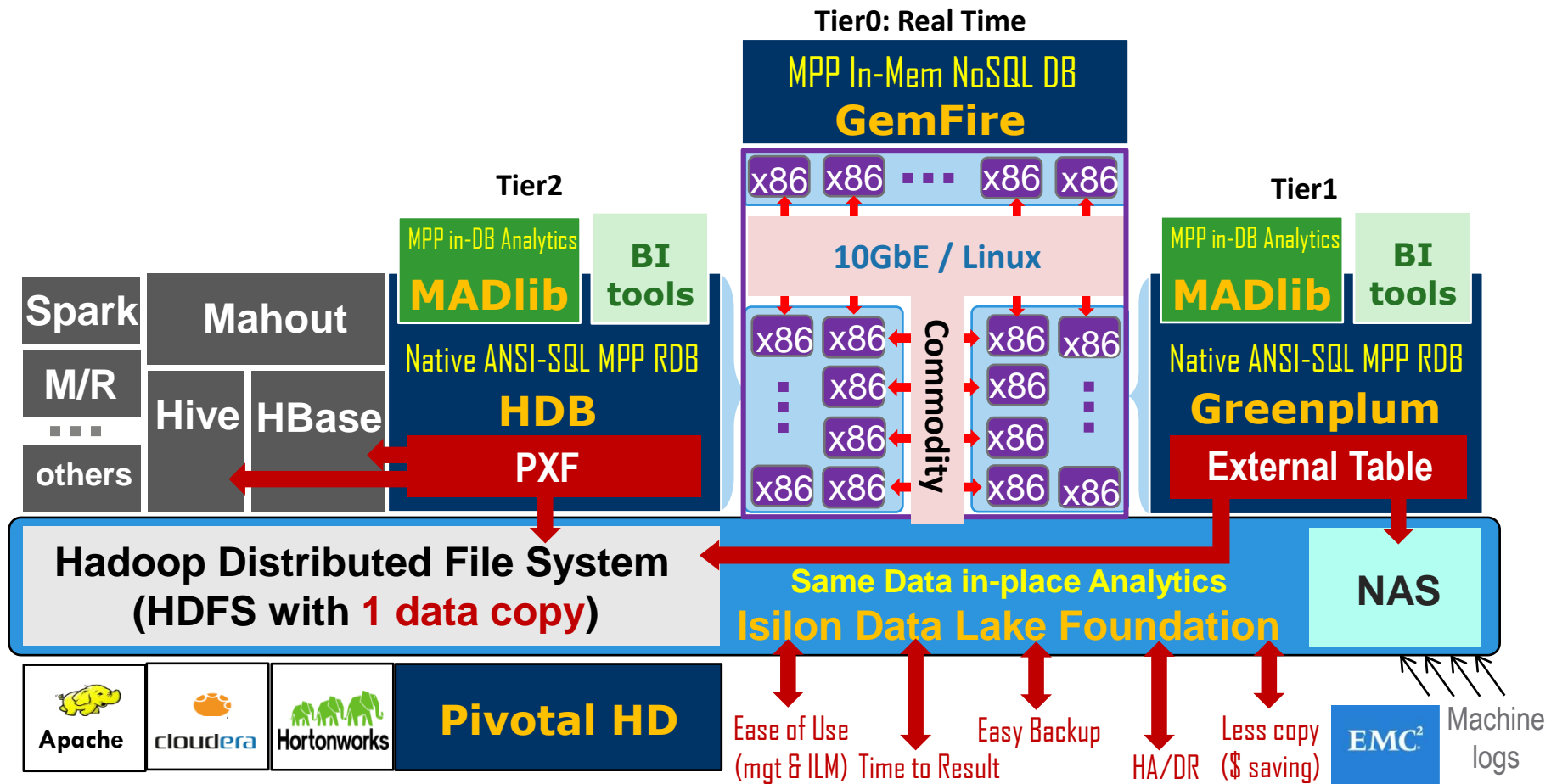
**Transaction-based
(In-Memory DB)**

**Search-based
(Hadoop)**

*** Source: IDC 2012 Big Data study

EMC²

X86+Isilon: 先進Hadoop技術架構+Pivotal BDS堆疊



加速創新關鍵: Pivotal BDS提供網格及大量平行運算架構 及免費的In-DB-computing先進演算法

MPP Structured, Unstructured, and In-Memory Grids

Integrated HDFS, massively parallel processing of Greenplum DB, HAWQ on Pivotal HD; Gemfire in-memory data grid for real-time intelligence

MPP Language, API, and Partner Integration

Functions and models in SAS, PL/R, C, PL/Java, PL/Perl, and PL/Python, **PostGIS**

Text Analytics

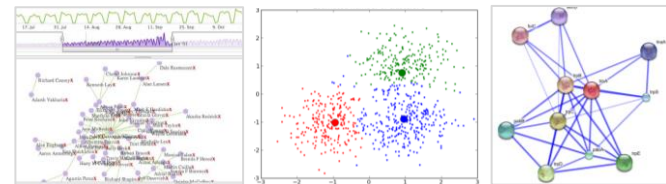
Built-in MPP text analytics of unstructured data
Faceted search and multi-lingual support
Semantic understanding through machine learning

Machine Learning

Open-source MPP library of advanced analytic functions including time series, linear and multinomial regression; supervised and unsupervised machine learning modules including SVM, LDA, and K-Means clustering

Graph Analysis

Open-source MPP library including Graph Analytics, Graphical Models, Clustering, Collaborative Filtering, and Topic Modeling



MADiib

- MAD** stands for:
 - M**AGNETIC
 - A**GLE
 - I**DB
- lib** stands for library of:
 - advanced (mathematical, statistical, machine learning)
 - parallel & scalable
 - in-database functions
- Mission:** to foster widespread development of scalable analytical skills, by harnessing efforts from commercial practice, academic research, and open-source development.

MADiib: A Open Community Project

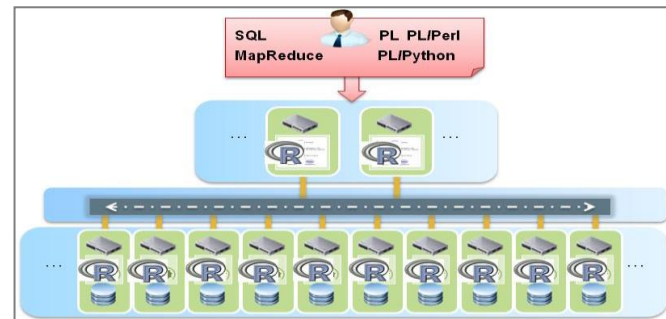
- Open Source: BSD license
- Developed as a partnership with multiple universities
 - University of California-Berkeley
 - University of Wisconsin-Madison
 - University of Florida
- Compatibility with Postgres and Greenplum Database.
- Designed for Data Scientists to provide Scalable, Robust Analytics capabilities for their business problems
- Homepage: <http://madlib.net>
- Source: <https://github.com/madlib>
- Forum: <http://groups.google.com/group/madlib-user-forum>

Community

open source

github

FREE

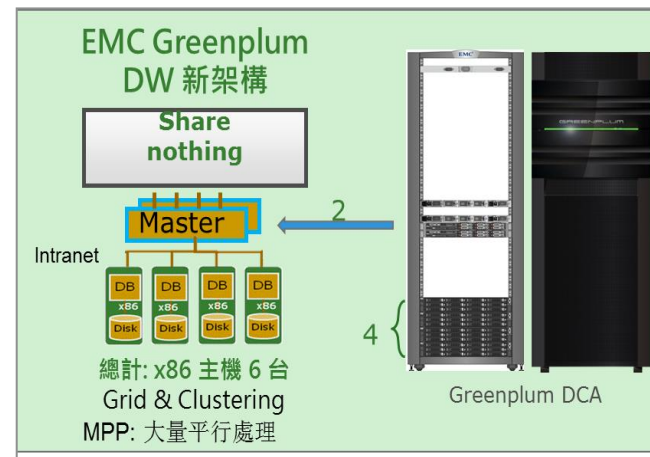


台灣高科技A公司 Greenplum EDW 應用案例

所有機台報表查詢測試(BIG TABLE)

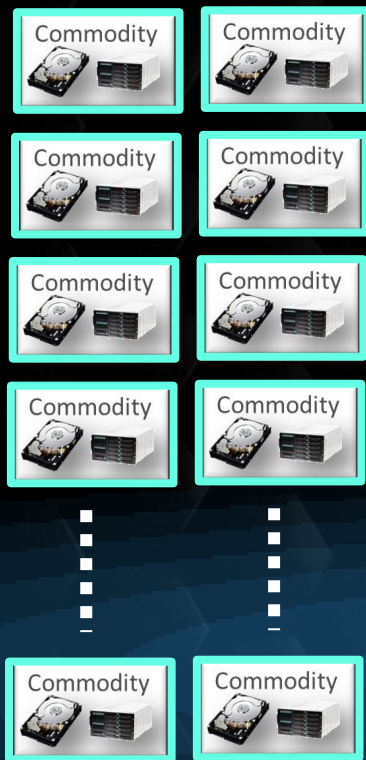
Greenplum DB					
CMTS Server:10.xx.yy.zz					
Win Server 2008					
Lot	Report Time(Sec)	DB Time(Sec)	Query Count	TABLE count	
XA3x01z4	8.5	1.7	17,334	15,779,604,540 ~ 3TB	
XA3x01z5	8.2	0.8	17,337		
XB3x01z3	7.5	1	17,340		
XB3x01z4	8.9	0.8	17,340		
XC3x01z5	7.4	1.1	17,328		
XC3x01z6	9	0.9	17,343		

TEST3: (BIG TABLE)
18 Table整合成1
個table測試Lot



小機器立大功, 透過Greenplum解決資料孤島及DB效能問題!

加速創新關鍵：在Edge端或Cold Data使用軟體定義儲存架構



**Multi-Protocol
Scale-Out File**

IsilonSD



**Cloud-Scale
Object**

ECS



**Hyper-Converged
Scale-Out SAN**

ScaleIO



Build Better Clouds, Modernize Apps, Analyze More Data,
Accelerate Performance

Management and Automation

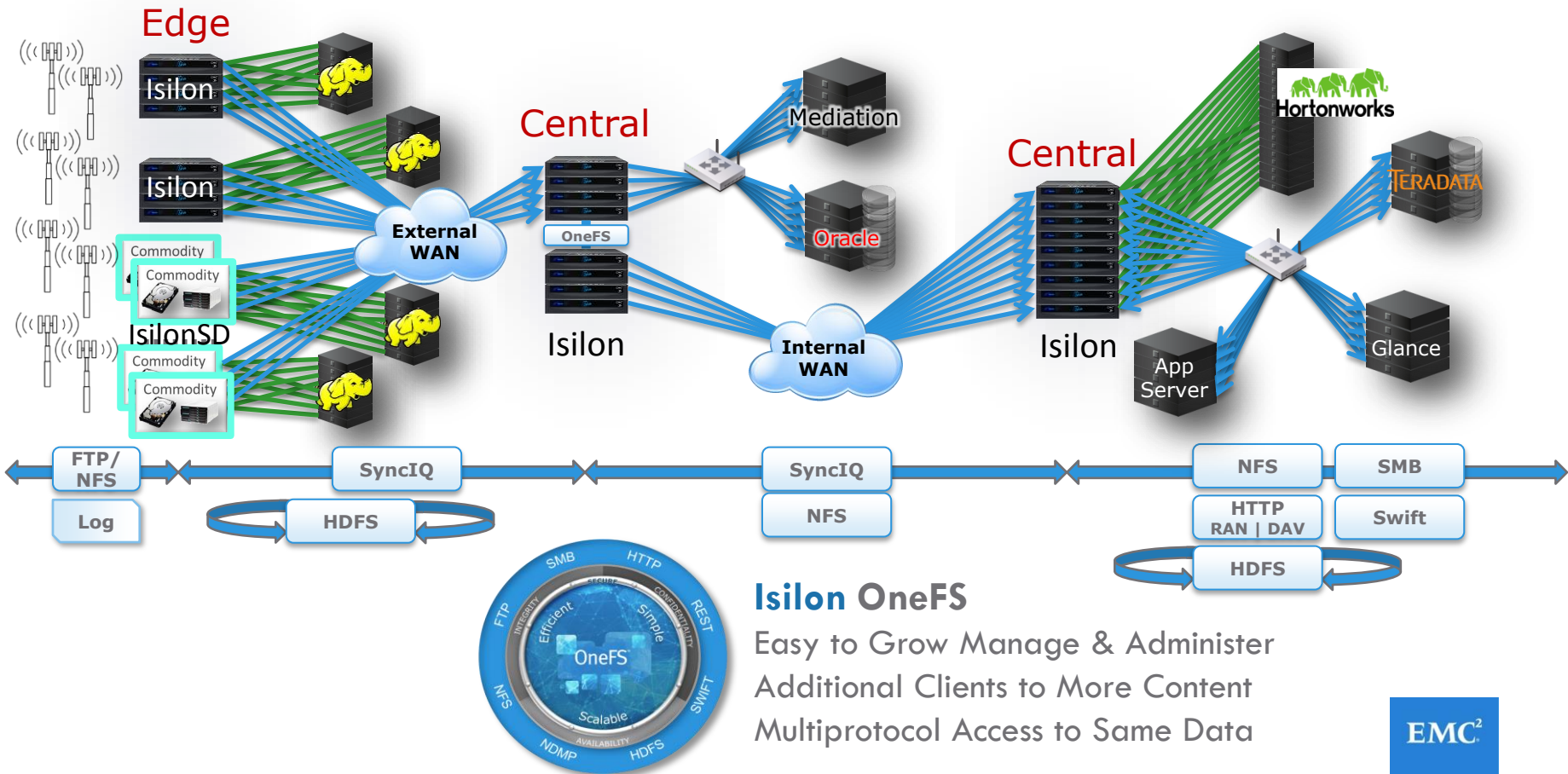
ViPR



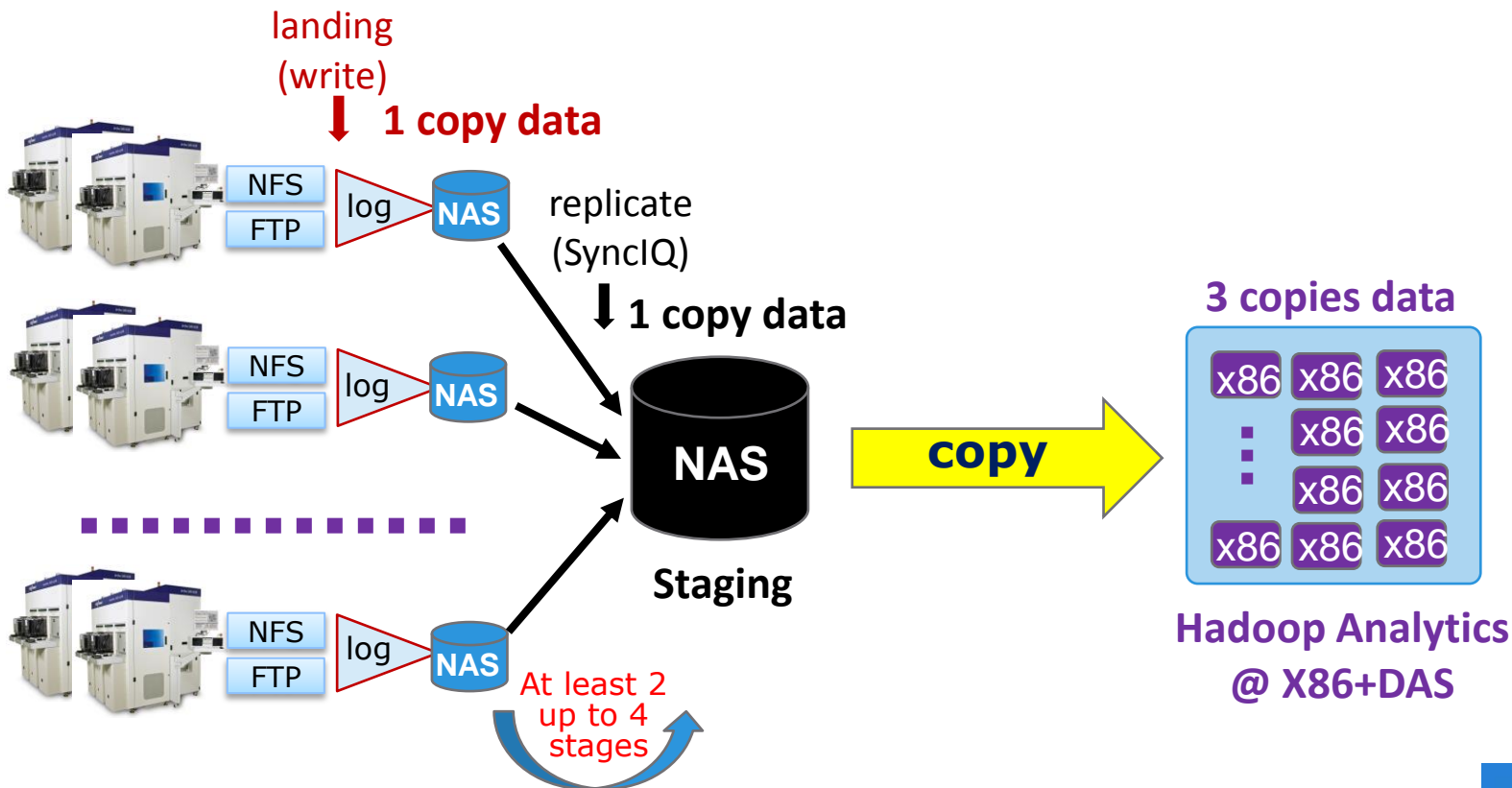
SUITE

Transform Multi-Vendor Storage

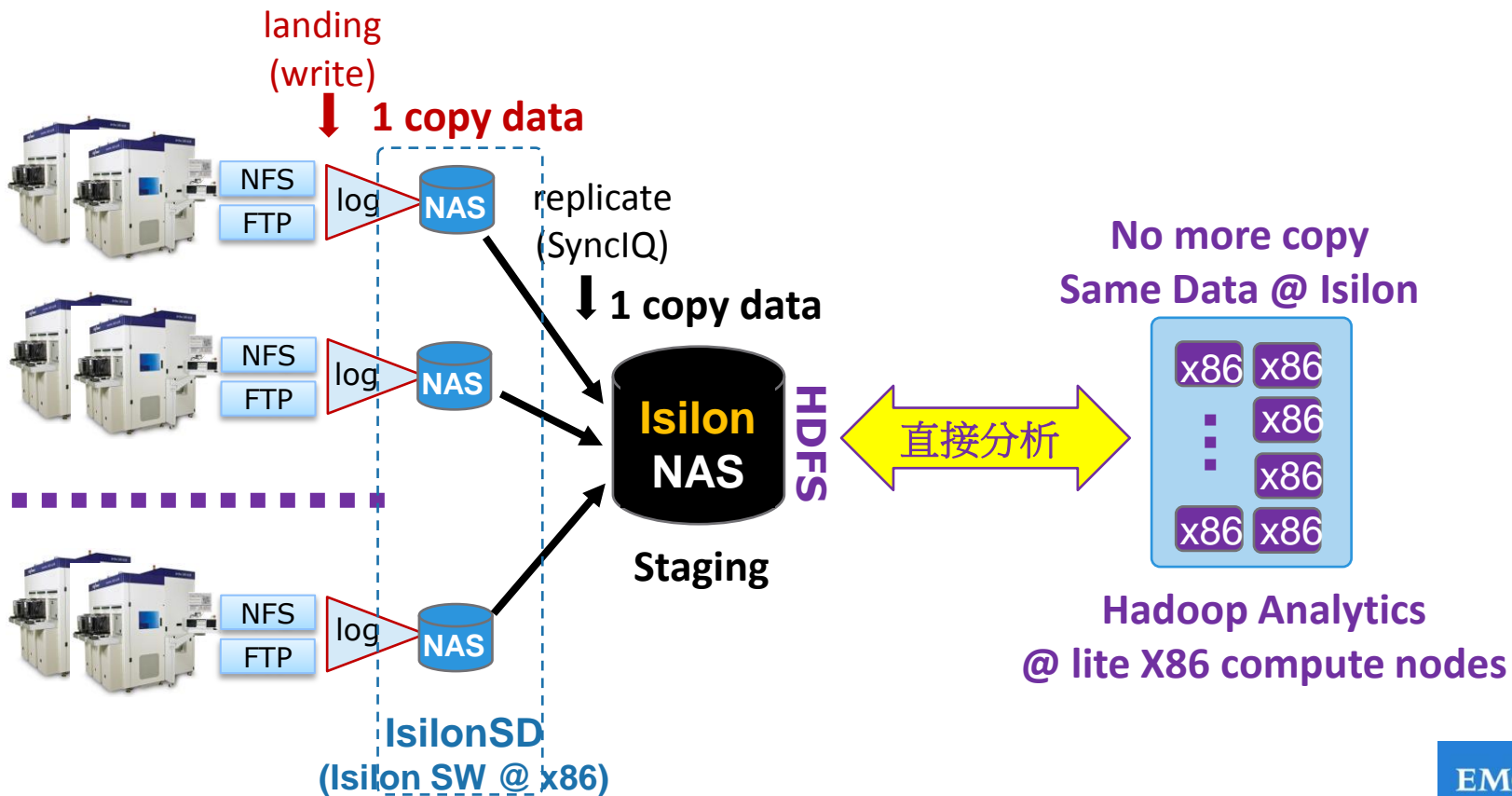
加速創新關鍵: 整合末端與中央分析應用流程(Isilon)



台灣高科技C公司既有Hadoop DAS架構與NAS資料流

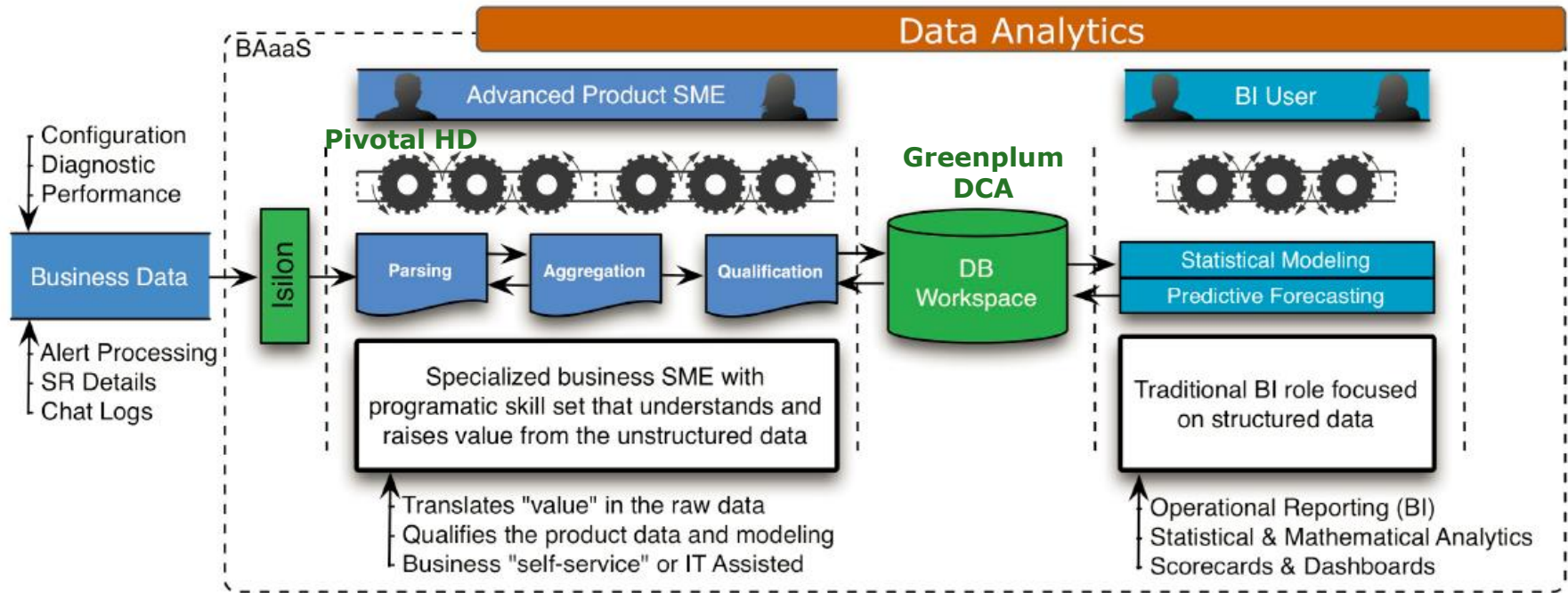


台灣高科技C公司新Hadoop+Isilon建議架構與資料流



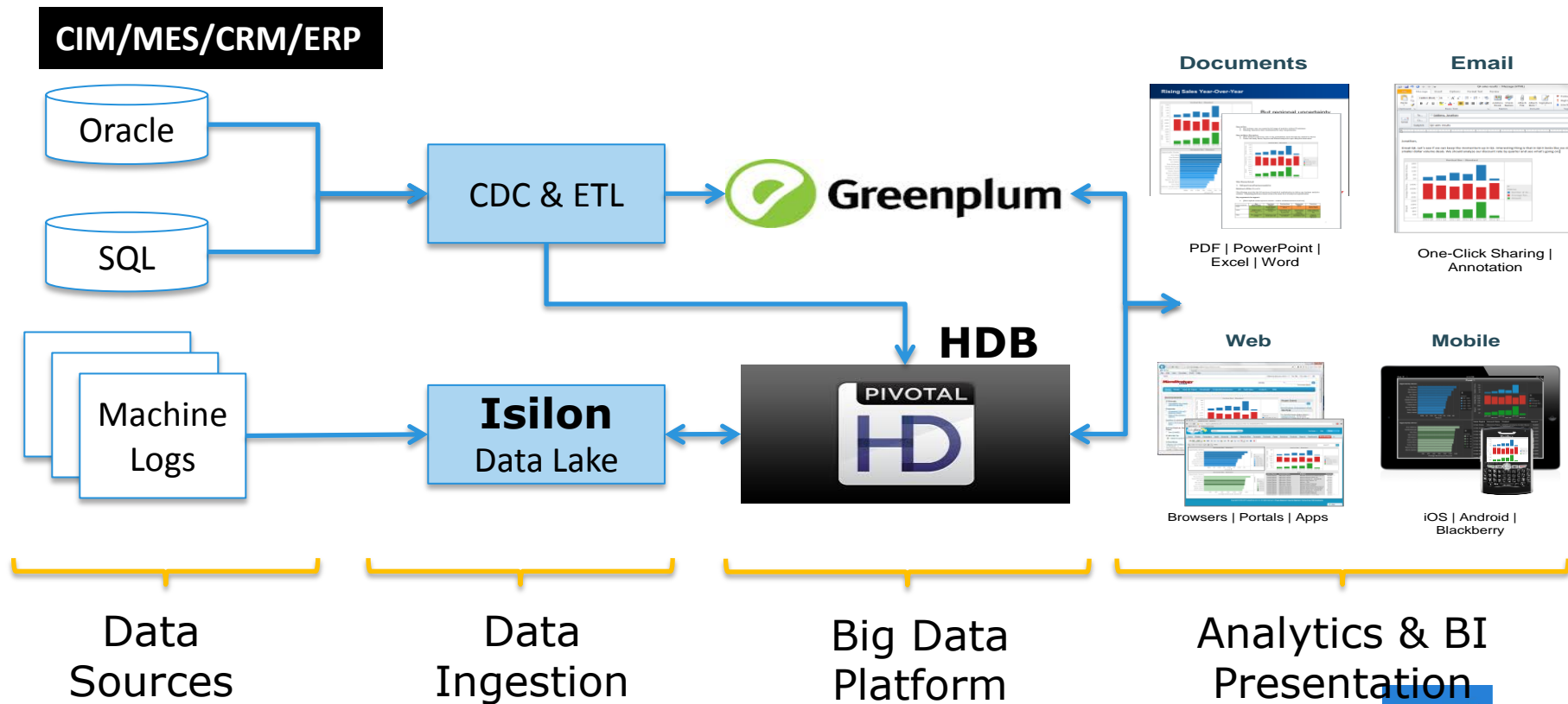
案例: EMC IT (Intranet IT & Machine logs Analytics)

Business Analytics-as-a-Service Engagement Process: Isilon (NAS, FTP, HDFS)+ PHD + DCA



IT enabled and Business driven for innovation, rapid delivery, and self-service

EMC Big Data 與傳統 DB 資料載入整合架構



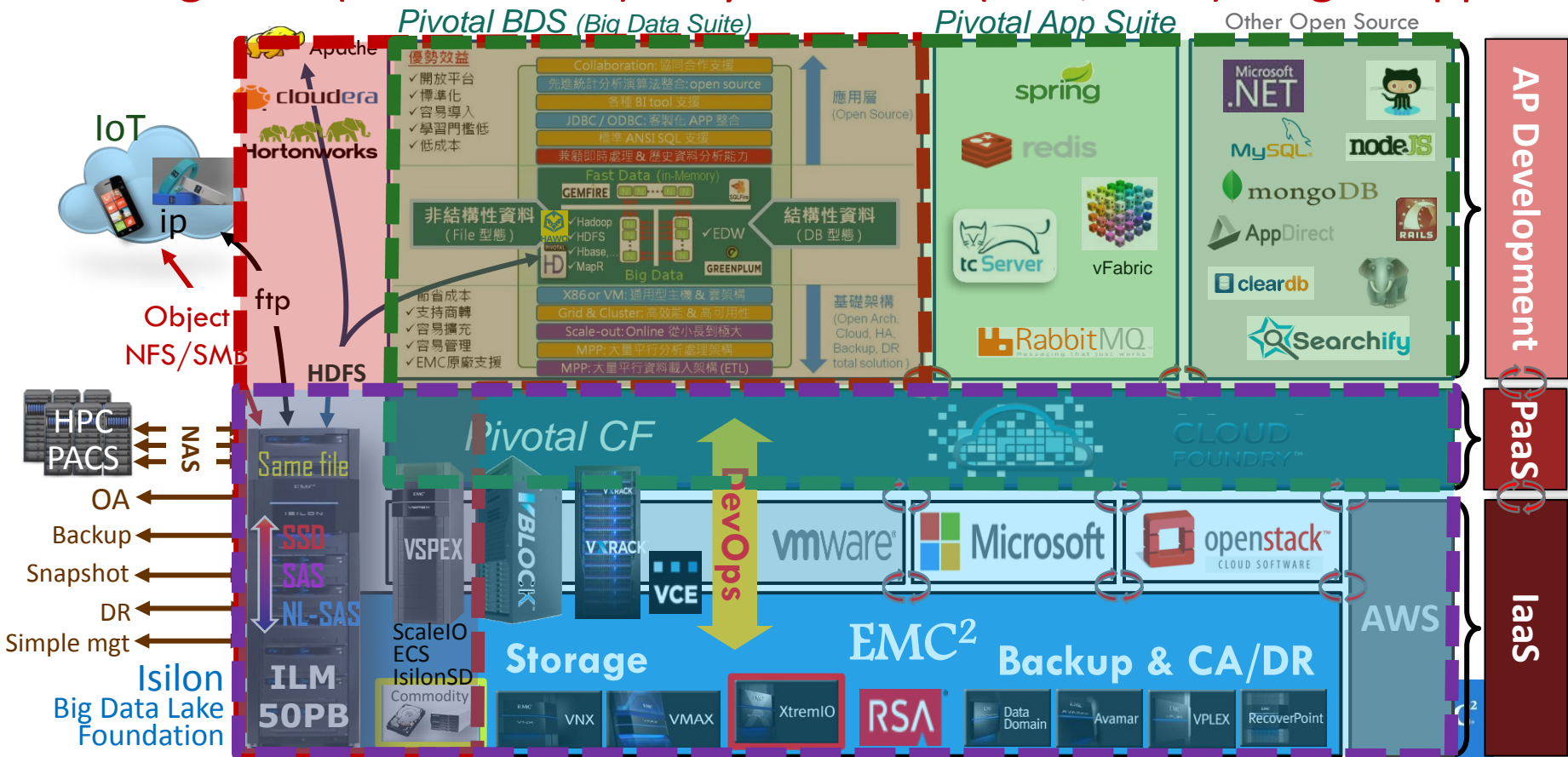
台灣高科技B公司Big Data分析平台使用性分類

Critical Application Priority

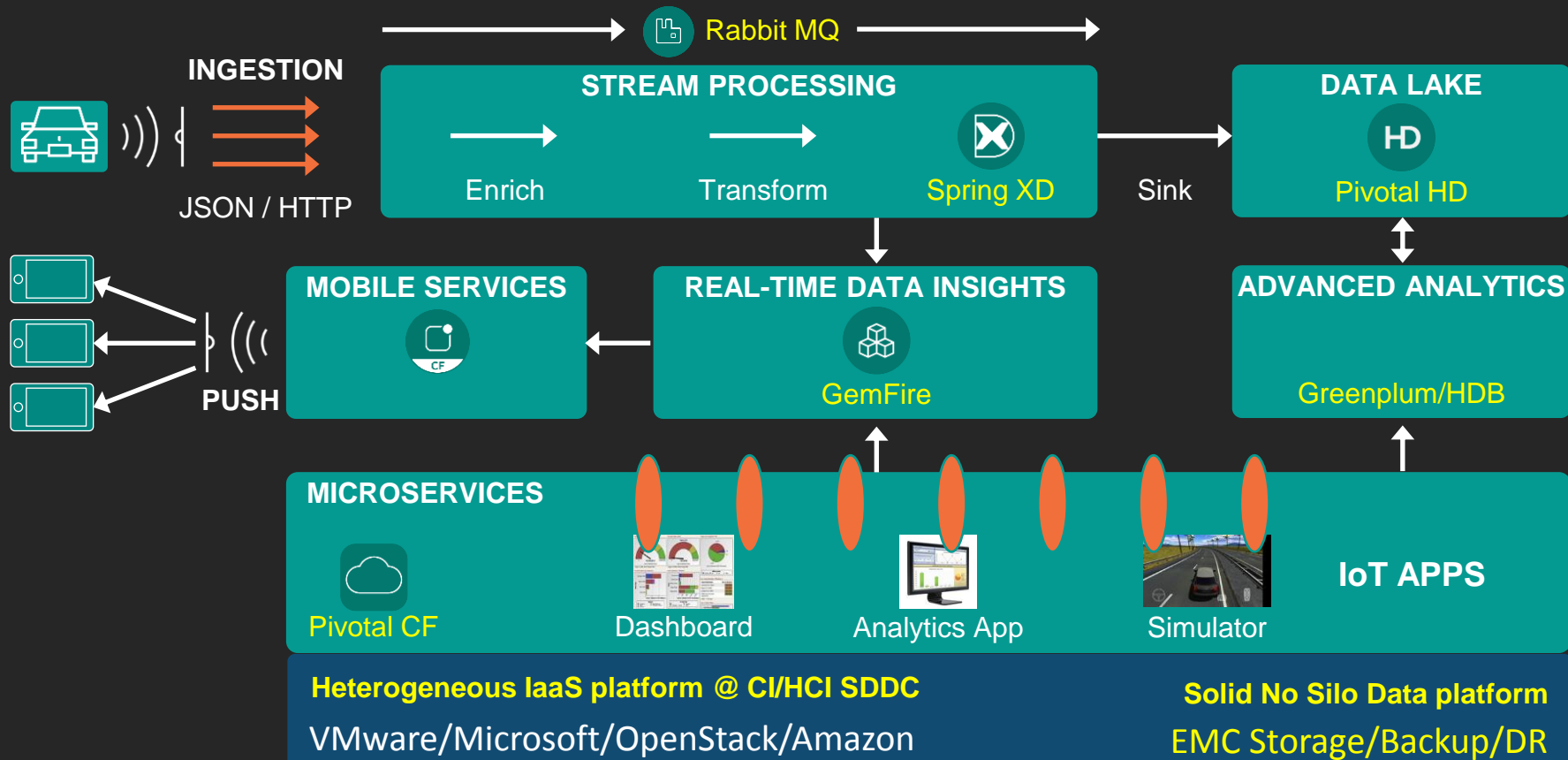
等級類別	系統名稱	可用空間	應用方向
Tier 1	EMC Greenplum DCA	9/36 TB	提供即時性統計分析需求
Tier 2	EMC Isilon + Pivotal HD & HDB	70 TB	提供高可用性、高運算能力及高擴展性資料平台
分析工具	MADlib	In-DB Analytics	數學、統計與機器學習等先進演算法

加速創新關鍵: 最佳工程整合新世代ITaaS架構 (engineered)

Big Data (No Data Silo) + Hybrid Cloud(IaaS, PaaS) + Agile App

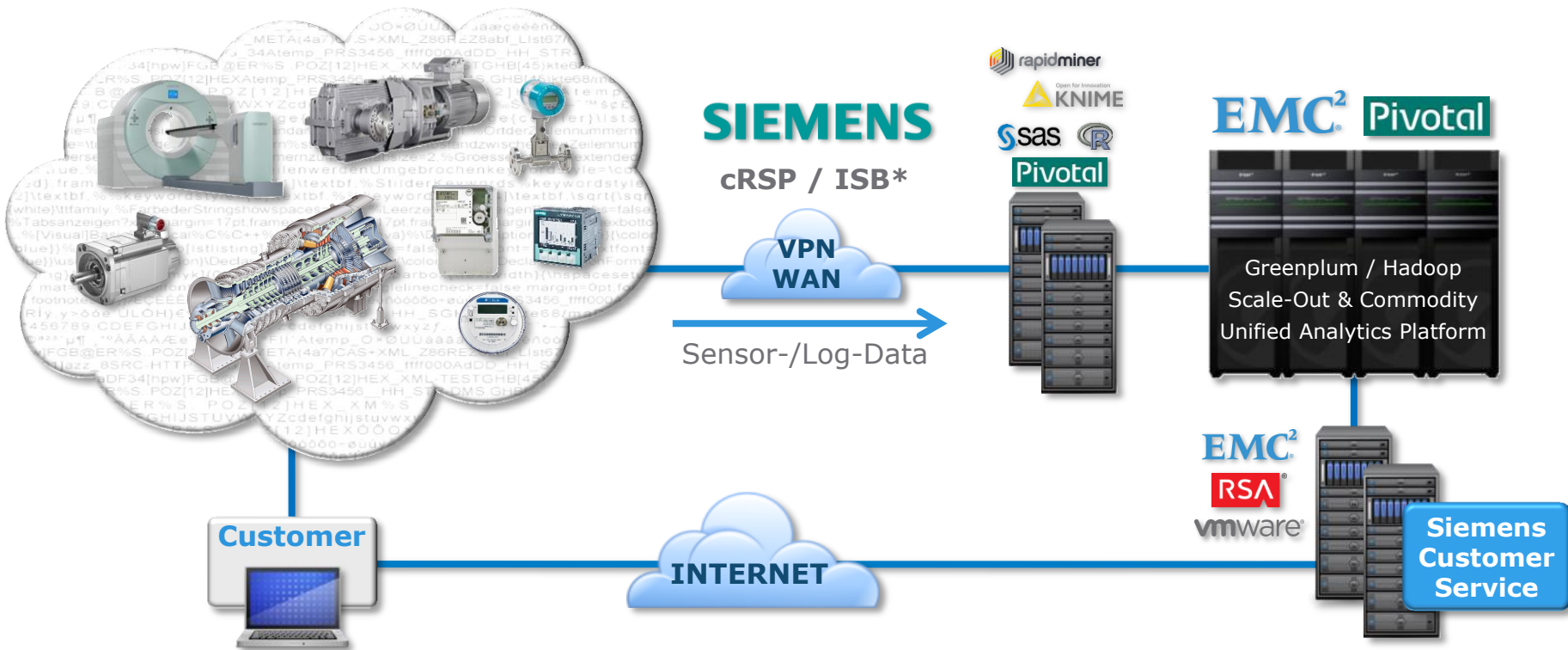


案例: EMC能將車聯網IoT中間所有片斷予以整合



案例：德國西門子 - 數位工廠基礎架構

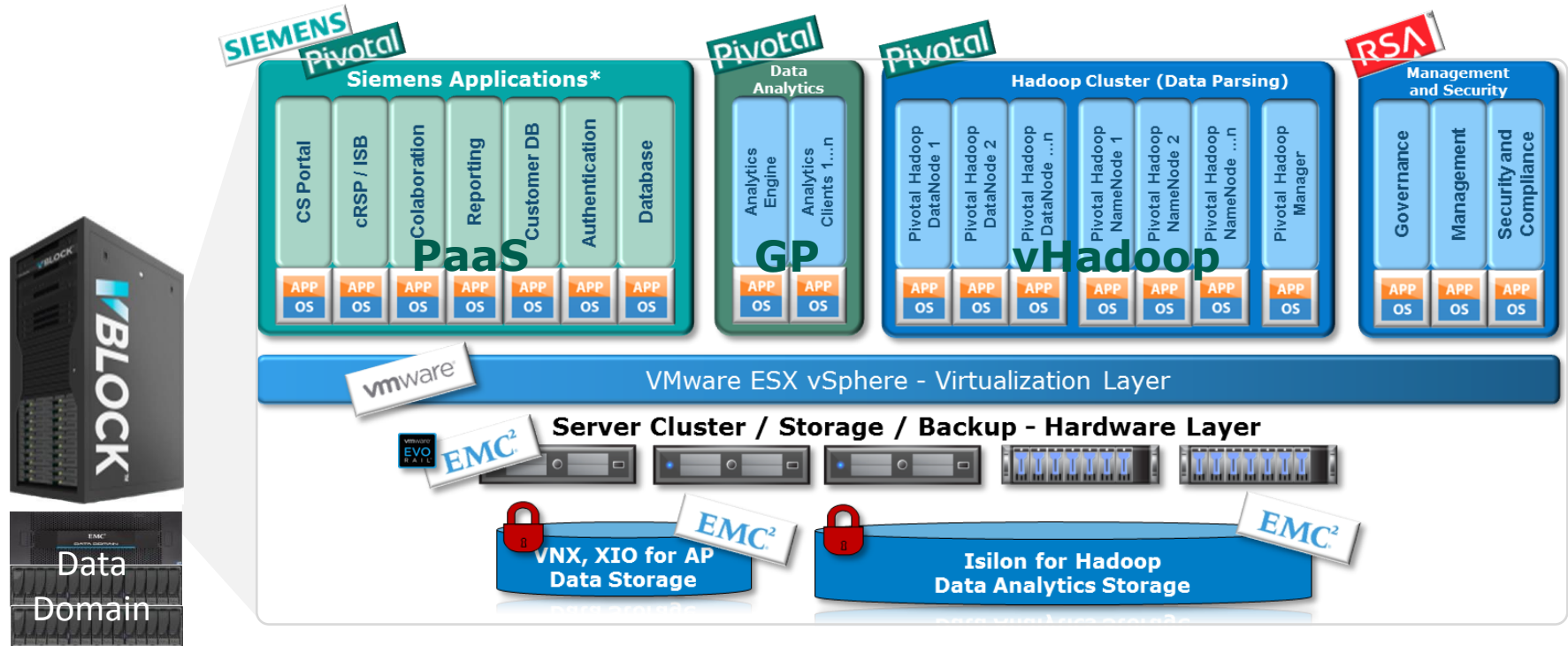
Smart Data for more Business Intelligence, Condition Monitoring and Predictive Maintenance



(*) cRSP = Common Remote Service Platform ISB = Industry Service Backbone

案例：德國西門子 - 數位工廠基礎架構

EMC Federation Cloud Perspective (Layer & Container) & Provider PaaS-Solution



= Encrypted Data Storage (optional)

(*) Sample Application Set. Depending on the usecase, the type and amount of applications can vary.



Big Data Project Keys to Success

MAXIMIZE OPPORTUNITIES

Current Situation

- Unclear business cases
- Skills deficits
- Lack of experience
- Rigid app dev procedures
- Complex app deployment
- Data silos
- Escalating data management costs



PEOPLE

PROCESS

TECHNOLOGY

Data-Driven Enterprise

- Optimal use cases
- Trained & experienced staff
- Agile dev methodology
- Technology: PaaS, Data Lake solutions
- Simplified data management

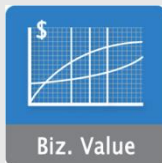
EMC²

People: EMC Big Data 課程

GAIN SKILLS FOR IMMEDIATE & EFFECTIVE PARTICIPATION IN BIG DATA PROJECTS

EXECUTIVES

90 min



Introducing Data Science & Big Data Analytics for Business Transformation

BUSINESS LEADERS

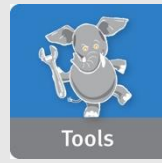
1 day



Data Science & Big Data Analytics for Business Transformation

DATA SCIENTISTS

5 days
(可客製化)



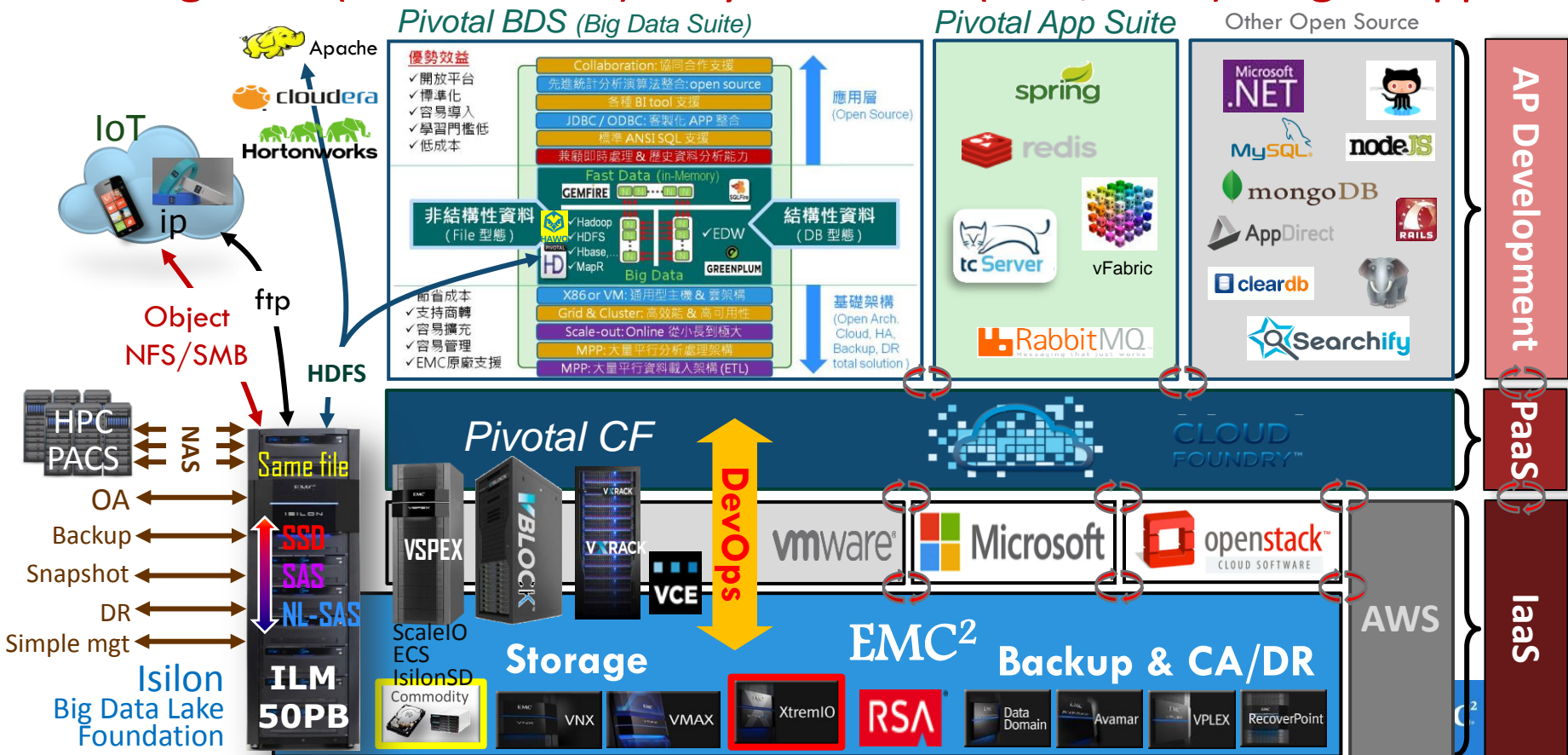
Data Science & Big Data Analytics

https://education.emc.com/guest/campaign/data_science.aspx

EMC²

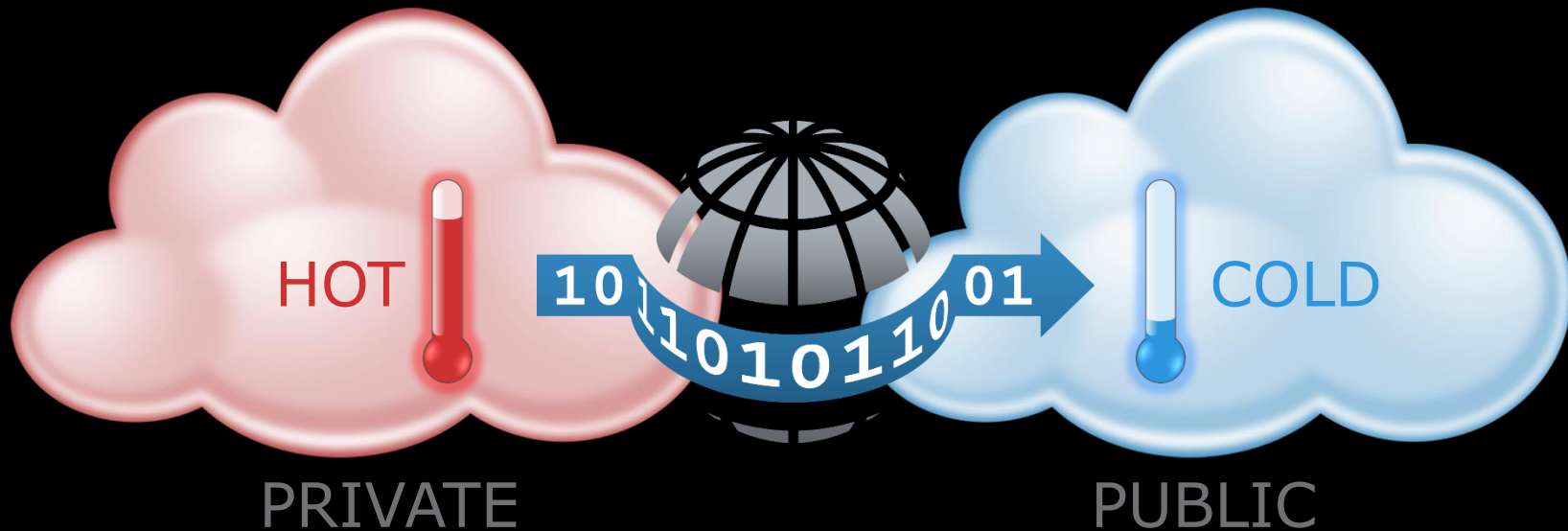
加速創新關鍵: 最佳工程整合新世代ITaaS架構 (engineered)

Big Data (No Data Silo) + Hybrid Cloud(IaaS, PaaS) + Agile App



混合雲應用

利用便宜的雲儲空間擴展資訊生命週期管理



EMC CloudArray是磁碟資料歸檔到雲端的利器



EMC CloudBoost是備份保護資料上雲端的利器

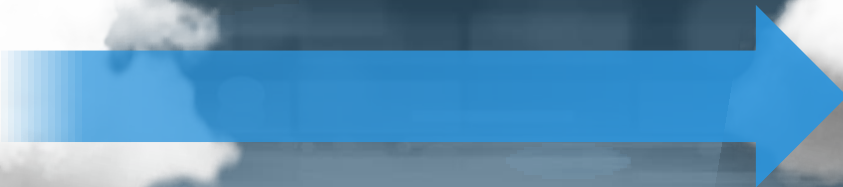


EMC²

EMC SPANNING是公有雲SaaS服務產生資料的備份利器



Office 365
salesforce
Google Apps



CLOUD TO CLOUD BACKUP



EMC²

EMC提供業界最完整直上雲端的儲存整合解決方案



SIMPLIFY &
AUTOMATE
STORAGE



PATH TO
THE
CLOUD



DATA
PROTECTION
EVERYWHERE



EMC²

EMC²®